DOCUMENT RESUME

ED 286 940 TM 870 607

TITLE	Computerized Adaptive Testing. ERIC Digest.
INSTITUTION	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	[83]
CONTRACT	400-83-0015
NOTE	4p.; ERIC/TME Update Series.
AVAILABLE FROM	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, NJ 08541-0001 (free).
PUB TYPE	Information Analyses - ERIC Information Analysis Products (071)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Adaptive Testing; *Computer Assisted Testing; *Item Banks; *Latent Trait Theory; Testing Problems; Test Items
IDENTIFIERS	ERIC Digests

ABSTRACT

This brief overview notes that an adaptive test differs from standardized achievement tests in that it does not consist of a certain set of items that are administered to a group of examinees. Instead, the test is individualized for each examinee. The items administered to the examinee are selected from a large pool of items on the basis of the correctness of the examinee's previous responses. Several steps are involved in constructing a computerized adaptive test (CAT): (1) the item pool must be developed; (2) some item characteristics must be calculated for each item; and (3) a strategy for deleting items from the item pool must be chosen, along with a scoring procedure. Latent trait theory assumes that an underlying trait is the sole determinant of an examinee's responses to test items, and that performance on an item is governed by two factors: the examinee's ability and the item's difficulty. Several item selection strategies have been developed, including the Robbins Monro procedure, fixed step size, the flexilevel method, the Bayesian approach, and the stratified-adaptive approach. Advantages of CAT are its increased efficiency, optimization of the difficulty level of the test for each examinee, and administration convenience. The most obvious constraint on the use of CAT is the problem of availability of appropriate computer hardware and software. (LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED286940

ERIC DIGESTS

COMPUTERIZED ADAPTIVE TESTING

ERIC Clearinghouse on Tests, Measurement,
and Evaluation, Educational Testing Service,
Princeton, New Jersey

1983

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM870607

COMPUTERIZED ADAPTIVE TESTING

An adaptive test is very different from standardized achievement tests. It does not consist of a certain set of items that are administered to a group of examinees. Instead, the test is individualized for each examinee. The items administered to the examinee are selected from a large pool of items on the basis of the correctness of the examinee's previous responses. To begin the process, an initial estimate of an examinee's ability level is obtained from previous test scores, grade level, or some other indicator. This estimated ability level is used to choose an item with an appropriate difficulty level from a pool of items. If the item is answered correctly, the next item selected is more difficult; if the item is answered incorrectly, the next item selected is easier. This process is continued until the examinee's true ability level is estimated with sufficient accuracy.

There are several steps in constructing a computerized adaptive test (CAT). First, the item pool must be developed. It must contain many items at varying difficulty levels, all measuring the same ability. Next, some item characteristics, called item parameters, must be calculated for each item. Then a strategy for selecting items from the item pool must be chosen, along with a scoring procedure. The choice of item selection strategy and scoring procedure is made by comparing the amount of information each selection strategy/scoring procedure combination provides about the examinee's ability.

CAT AND LATENT TRAIT THEORY

Latent trait theory assumes that an underlying trait is the sole determinant of an examinee's responses to test items; and that level of performance on an item is governed by two factors: the examinee's ability and the difficulty of the item. More specifically, it is assumed that for any given item, a person with a higher ability should have a higher chance of passing an item than a person with a lower ability; and for persons with the same level of ability, the chance of passing an easier item should be higher than the chance of passing a more difficult item. These assumptions can be expressed mathematically.

Latent trait theory has certain advantages over classical test theory when applied to CAT. First, latent trait theory allows all the test scores to be expressed on one scale, regardless of the test administered. This allows the easy comparison of examinees. Second, in the mathematical models developed as expressions of latent trait theory, the test item parameters can be calculated independent of the group taking the test, rather than being tied to a certain sample of examinees.

Several different mathematical models have been developed to express the assumptions of latent trait theory. These latent trait models take into account different item parameters. All the models use difficulty level as an item parameter. In addition, some models take into account the discriminating power of an item and the guessing behavior of the examinees.

ITEM SELECTION STRATEGIES

Several item selection strategies have been developed. In each of them, a more difficult item is selected if the examinee responds correctly; an easier item is selected if the examinee responds incorrectly. Three of the strategies differ in the size of the difference in difficulty level between the last item selected and the next item selected. These three strategies are:

Robbins Monro Procedure. The amount of difference in item difficulty between the current item and the next item gradually decreases, so that the item difficulty converges on the level that is equal to the examinee's ability level.

Fixed Step Size. The amount of difference in item difficulty is kept constant. The item difficulty levels do not converge; they vascillate between just above and just below the examinee's ability level.

Flexilevel. This method uses a smaller item pool than the previous two methods. After an item is answered correctly, the next harder item is administered. After the item which is equal to the examinee's true ability level, the selected items will oscillate between those that are much easier and those that are much harder.

TM 870 607

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, AND EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NJ 08541-0001 (609) 734-5181

ERIC
TESTS MEASUREMENT EVALUATION

Two more strategies are also commonly used. With the Bayesian approach, the examinee's ability level is estimated from each response. Items are selected such that the difference between item difficulty and the estimated ability is the minimum possible. In the stratified-adaptive, or stradaptive, approach, items are arranged into strata by difficulty, then within a stratum by discriminating power. Item selection moves from stratum to stratum, starting with the most discriminating item in each level.

ADVANTAGES OF CAT

Computerized adaptive testing has several advantages over group-administered standardized tests. A major advantage of CAT is its increased efficiency. An accurate estimate of an examinee's ability can be obtained by administering fewer items. This is advantageous both for the examiner and the examinee.

Second, an adaptive test can individually optimize the difficulty level of the test for each examinee. This allows the test to do a better job of measuring the ability levels of individuals, especially at the extremes of ability; and it allows the examinee to experience much less frustration from responding to items which are much too difficult or too easy.

The test administration setting for CAT also provides several advantages. The test can be scheduled at the examinee's convenience. Test security is increased because no test booklets are printed and each examinee's "test" is different. The administration of the test is more standardized because all administration is handled by the computer. The test results can be made immediately available to the examinee. Finally, additional information about the examinee can be collected, such as the amount of time taken to respond to specific items or the consistency of responses across items.

CONSTRAINTS TO USING CAT

The most obvious constraint on the use of CAT is the availability of appropriate computer hardware and software. Adequate storage space for large item pools and item parameter data must be available, along with programs that will respond quickly enough to keep examinee waiting time at an acceptable level.

The item pool itself must be developed and must be checked to make sure that all the items in the item pool are measuring the same ability. If they do not, latent trait theory is not appropriate, and the test will not provide an accurate estimate of the examinee's ability.

Finally, it has been found that, using an item pool of 500 items, you actually get better accuracy by administering the 60 most discriminating items, rather than using adaptive testing techniques to select the item at the most appropriate difficulty level. This disadvantage must be overcome to make CAT truly practical.

REFERENCES

1. Gustafsson, Jan-Eric. An Introduction to Rasch's Measurement Model. ERIC/TM Report 79. Princeton, N.J.: ERIC Clearinghouse on Tests, Measurement, and Evaluation, September 1981. (ERIC Document Reproduction Service No. ED 211 594).

2. Kreitzberg, Charles B.; And Others. Computerized Adaptive Testing: Principles and Directions. Computers & Education 2(4): 319-29, 1978. (ERIC No. EJ 196 505).

3. Reckase, Mark D. Tailored Testing, Measurement Problems and Latent Trait Theory. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Los Angeles, CA, April 14-16, 1981. (ERIC Document Reproduction Service No. ED 207 987).

4. Rudner, Lawrence M. A Short and Simple Introduction to Tailored Testing. Paper presented at the Annual Meeting of the Eastern Educational Research Association, Williamsburg, VA, March, 1978. (ERIC Document Reproduction Service No. ED 166 257).

5. Wainer, Howard. On Item Response Theory and Computerized Adaptive Tests: The Coming Technological Revolution in Testing. Journal of College Admissions 28(4): 9-16, April 1983.

A more complete bibliography on "Computer-Assisted Tailored Testing" is available for $6.00 from the ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, NJ 08541-0001.

The National Institute of Education

This publication was prepared with funding from the National Institute of Education, U.S. Department of Education under contract no. NIE-400-83-0015. The opinions expressed in this report do not necessarily reflect the positions or policies of NIE or the Department of Education.